

全期間及び短期間双方の生体信号の使用による 長期ストレスレベル認識精度の向上 Improvement in Chronic Stress Level Recognition by Using Both Full-term and Short-term Measurements of Physiological Features

中島 嘉樹*¹
Yoshiki Nakashima

辻川 剛範*¹
Masanori Tsujikawa

大西 祥史*¹
Yoshifumi Onishi

*¹ 日本電気株式会社 データサイエンス研究所
Data Science Research Laboratories, NEC Corporation

In this paper, we propose improvement in chronic stress level recognition by using both full-term and short-term physiological features. In our proposed method, we employ the characteristic of PSS (Perceived Stress Scale), a widely used chronic stress measure. PSS scores are known to be influenced by mental states caused by high-stress experiences that could occur over shorter terms. So we added new stress features calculated on a weekly basis to the conventional stress features calculated on a monthly basis for which PSS questionnaires recognizes stress level. With weekly-basis feature calculations, we are able to recognize high-stress experiences over shorter terms. To evaluate our proposed method, we performed experiments using a 33-employee, 1-month database of physiological signals. Results have shown the Pearson's correlation coefficient to improve from 0.66 to 0.72 with use of our method.

1. はじめに

職業性ストレスは、鬱病等の精神疾患の発症をもたらし、生産性の低下や離職・休職等の原因となっており、被雇用者にとって大きな問題である。雇用者にとっても、生産性や労働力の減少は対策コストを強いられる課題である。ストレスを早期に見出し、それを防止することは、被雇用者・雇用者双方に大きな利益がある。

ストレスには、短期ストレスと長期ストレスがある。短期ストレスとは、身体的または心理的ストレスへの短期的な反応を指す。ストレスへの短期的な反応は、生体信号を用いて容易に検出できるため、Alberdi らがまとめているように[1]、生体信号からストレスへの短期的反応の程度を認識する多くの先行研究がある。それに対し、長期ストレスは、長期に亘ってストレスに曝されることにより起こるもので、鬱病等の原因となるため、より深刻な問題であるが、長期間に亘る継続的な測定が必要であるため、研究の難度は高く、先行研究は少ない。

生体信号を用いた長期ストレスレベル認識の重要な先行研究の一つは、Sano ら[2][3]によって報告されたものである。この先行研究では、68人 [2]または165人 [3]の学生の1ヶ月分という、比較的大きなデータベースを使用しており、当該データベースには、生体信号としてEDA(皮膚電気活動)、ST(皮膚温度)、および加速度計信号が含まれている。当該先行研究では、心理学的に信頼できる長期ストレスアンケートであるPSS(Perceived Stress Scale)[4]のスコアを元に2クラス(PSSスコア上位20%、下位20%の2クラス)を定義し、その2クラスの分類を行っている。

しかし、ストレスレベルの全範囲を認識するのに有効な技術を開発するためには、ここで省略された中位60%も考慮する必要がある。本稿では、特にこの中位ストレスに対するストレスレベル認識精度の改善について、全期間およ

び短期間の双方で算出した生体信号特徴量を用いて改善することを提案する。我々は広く使われている長期ストレス尺度であるPSSを採用することとした。PSSは一ヶ月単位での長期ストレスへの曝露の程度を認識するだけでなく、一ヶ月より短い期間で起こり得る強いストレスへの曝露によって引き起こされる心理状態の悪化(心理的不調)も反映する特性が知られている[5]。また我々の考察ではこの特性は特に中位のストレスレベルにおいて顕著に表れる。よって、PSSによるストレスレベルをより精度よく認識する為に、PSSのこのような特性も考慮し、我々は、1ヶ月単位だけでなく、より短い期間である1週間単位でもストレス特徴量を算出した。これによって、中位を含む全範囲のストレスレベルを精度よく認識することができる。

以下、本稿の構成について述べる。第2節では、Sanoら[2][3]の研究を先行研究として紹介し、その概要と課題を述べる。第3節では、我々が提案する方法について述べる。第4節では、提案法の実験的評価について述べる。第5節では、研究を要約する。

2. 先行研究

2.1 先行研究の概要

本節では、まず、Sanoらの先行研究[2][3]の概要について述べる。当該先行研究に代表される長期ストレスレベル認識システムは、概ね図1のような構成をとる。まず、被験者が着用するウェアラブルセンサ等を使用して、生体信号を取得する。たとえば、Sanoらの研究では、EDA、ST、加速度を取得している。これらは、それぞれ発汗、体温、

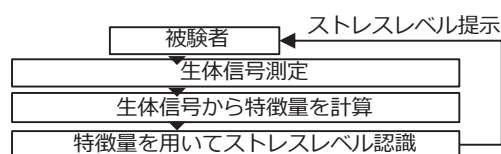


図1 ストレスレベル認識システム

連絡先: 中島嘉樹, 日本電気株式会社, 〒211-8666 川崎市
中原区下沼部 1753, 電話番号: 044-431-7554, 電子メ
イルアドレス: y-nakashima@bu.jp.nec.com

体動を測定する為のものである。次に特徴量の計算を行う。先行研究では、取得した生体信号から1ヶ月単位で一括して特徴量計算を実施している。このような1ヶ月単位の特徴量の例として、以下に示す2つの式(式(1)および(2))による、特徴 $a6G_M$ で示されるような3軸加速度計信号に基づく値が挙げられる。

$$x_i = \sqrt{x_{1,i}^2 + x_{2,i}^2 + x_{3,i}^2} \quad (1)$$

$$a6G_M = \frac{1}{N} \sum_{i=1}^N a6G_i, \quad a6G_i = \begin{cases} 0 & (x_i < 6G, 7G \leq x_i) \\ 1 & (6G \leq x_i < 7G) \end{cases} \quad (2)$$

ここで、式(1)において、 i は信号の系列番号であり、1から N の値を取る。 N は信号の総数である。 SR をサンプリングレート、 D を一月の合計日数とすると、 N は $D * 24 * 3600 * SR$ となる。 D は28, 29, 30, 31のうちどれかの値を取る。 SR は、先行研究においては8Hzであった。 $x_{1,i}$, $x_{2,i}$ および $x_{3,i}$ は、ウェアラブルセンサによって測定される、互いに直交する3軸の加速度である。測定された加速度信号は、 G すなわち重力加速度 $9.8m/s^2$ の単位で記録される。ここで、式(1)における x_i は3軸加速度のRMS(Root Mean Square)である。式(2)において、 $a6G_M$ は、全加速度計信号のうち、RMSが $6G \sim 7G$ の範囲にある信号が全体に占める割合を示す。 $a6G_M$ 等の、月単位での特定の範囲の加速度信号の全体に対する割合を計算した特徴量は、当該月の体動を反映する特徴量の一種である。小さい加速度の割合が大きい場合、その月の体動は落ち着いており、一方、大きな加速度の割合が大きい場合、その月の体動は激しいと言える。先行研究では、長期ストレスはこうした月単位の特徴量を用いて認識された。当該先行研究では、PSSスコア上位20%および下位20%のPSSスコアの被験者が、それぞれ高ストレスおよび低ストレスクラスとして定義され、2クラス分類で識別された。中位60%の被験者を省略し、ストレスレベルが極端に離れた2クラスのデータのみを用いて2クラス分類を行うことは、そこで用いられる式(2)のような長期ストレス特徴量を評価するには有効であった。

2.2 先行研究の課題

先行研究で開発された特徴量は、被験者を高ストレス・低ストレスの2クラスに分類するには有効であるが、PSSを長期ストレスの指標とする場合、特に中程度のストレスに対しては十分に有効ではない。PSSが測定するストレスの指標である「ストレス評価」は回答者が曝されたストレスラーの程度だけでなく、ストレスラーによって影響を受けた回答者の心理状態も同時に反映する特性がある為である[5]。PSSは一ヶ月間の「ストレス評価」を認識の対象とするが、回答者の心理状態は、一ヶ月より短い期間に起こり得る強いストレスラーによって悪化することがある。そして、回答者の「ストレス評価」は、たとえ回答者が当該月のその他の期間には弱いストレスラーにしか曝されていない場合でも、比較的高い値となる傾向がある。ところが、先行研究の式(2)のような月単位の特徴量では、このようなPSSの特性は反映されない。月単位での特徴量にPSSの特性が反映されない状況は、次章で詳細に述べるとおり、当該月のストレスラーに強弱の変動がある場合、つまり月単位で見れば中程度のストレスレベルの場合に顕著になる。

我々は、PSS、即ち「ストレス評価」がこのように回答者の心理状態の悪化の影響を受けるという特性は、リスクの高い回答者を識別するのにより有効であると判断し、そ

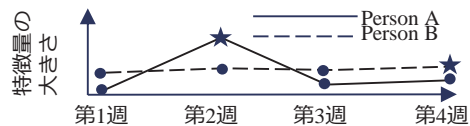


図2 提案特徴量のコンセプト

のようなPSSの特性を表現できる特徴量を開発する必要があると考えた。

3. 提案法

本節では、まず、前節で述べた先行研究の課題をより具体的に説明し、次にその課題を解決する為の提案法について述べる。まず、図2を参照する。図2には、式(3)で示される週単位特徴量のプロットが2つ示されている。具体的には、式(3)の $a6G_W_j$ を、 $j=1,2,3,4$ に亘ってプロットしている。式(3)は月単位の特徴量である式(2)の $a6G_M$ を週単位にしたものであり、月単位ではなく週単位のストレスレベルを反映できる特徴量になっている。

$$a6G_W_j = \frac{4}{N} \sum_{i=\frac{N}{4}(j-1)+1}^{\frac{N}{4}j} a6G_i \quad (3)$$

図2の2つのプロットのうち、実線、つまり「Person A」のプロットは、低いストレスレベルと高いストレスレベルの両方を含み、一方、破線、つまり「Person B」のプロットは、4週間全体に亘って中程度のストレスレベルを示している。前述のメカニズムにより、PSSスコアにおいては「Person A」が高くなり、実際リスクも「Person A」の方が高くなると思われるが、式(2)を使って月単位の特徴量を計算すれば、両者のストレスレベルは同程度になる。そのため、式(2)のような月単位の特徴量だけでなく週単位の特徴量を計算し、この週単位の特徴量プロットから統計値を取得することが重要であると我々は考えた。ここでいう統計値とは、週単位の特徴量プロットにおいて特定の週の高いストレス経験を的確に反映すると期待できる最大値、標準偏差等を意味する。例えば、星マークとして図2に示されている各プロットの最大値は、式(4)を用いて計算することができる。

$$a6G_W_{max} = \max\{[a6G_W_j]_{1 \leq j \leq 4}\} \quad (4)$$

式(4)のような特徴量を用いれば、式(2)では区別できなかった「Person A」と「Person B」の違いが明確になる。この効果は主に中程度の範囲内におけるストレスの高低の識別に有効であることには注意が必要である。ストレスレベルのプロットが連続的に高い場合、或いは連続的に低い場合には、式(2)と式(4)はほぼ同じ値を示すからである。

4. 実験

4.1 データ収集

実験の被験者は全て自発的に実験に参加した健康なオフィスワーカーであった。実験に参加することに同意する前に、実験の目的及び手順を説明した。33人が実験に参加した。33人のうち4人が女性であり、年齢は、20代が8人、30代が10人、40代が10人、50代が5人であった。被験者は

勤務時間中、毎日ウェアラブルセンサ (Empatica 社 E4 センサ [6]) を着用するよう要請され、センサによって取得された被験者の生体信号は毎日アップロードされた。被験者はオフィスから離れたときにはセンサは着用しなかった。センサがデータを取得した平均日数は 17 日であった。E4 センサは、4Hz のサンプリングレートで EDA と ST、32Hz で加速度信号を取得する。長期ストレスのアンケートとして PSS アンケートの 10 項目版 (PSS10) を使用した。1 ヶ月の実験期間の前後に PSS10 に回答するよう被験者に要請した。実験後の PSS アンケート回答に基づくスコアを長期ストレスの指標とした。図 3 がその分布を示すヒストグラムである。平均は 17.2、標準偏差は 5.7 であり、これは典型的な分布と考えられる [7]。

4.2 実験条件

1) 特徴量計算

先行研究 [2] を参照し、特徴量を検討した。表 1 に、本実験で使用した特徴量を示す。先行研究 [2] の特徴量計算における特徴的な手法の一つは、加速度信号を用いて活動状態 (座位、歩行、走行) を区別し、各活動状態のデータ毎に特徴量を計算することである。生体信号はストレスのような精神活動だけでなく、身体活動の影響も強く受けるので、活動状態をいくつかのカテゴリーに区別することは精神状態の影響をよりの確に把握する上で好ましい。我々は先行研究 [2] と同様、このアプローチを EDA と加速度計信号に適用し、更に、これも先行研究と同様、個々の活動状態の時間的割合をも特徴量とした。尚、表 1 の特徴量は全て、月単位と週単位の双方で計算可能である。

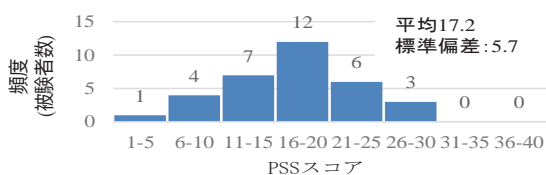


図 3 取得したデータベースにおける PSS スコアの分布

表 1 特徴量とその合計数

活動状態	モダリティ	特徴量 (頻度特徴量は、各ビン の頻度の合計も特徴量を含む)	特徴量 数	合計 特徴量数
座位、 歩行、 走行、 全体	規格化 EDA	平均	1	91
		中央値	1	
		標準偏差	1	
		周波数 (11ビン) (0.1Hz幅ビン、0~1.1Hz)	11+1	
		パワースペクトル密度 (5ビン) (0.1Hz幅ビン、0.0-0.5Hz)	5+1	
		EDA ピーク	30秒区間中のピーク数の 頻度 (11ビン)	
	EDA ストーム	1秒区間中のピーク数の 頻度 (11ビン)	11+1	
		30秒区間中のピーク が占める割合	1	
		30秒区間中の ピーク数のエントロピー	1	
	加速度 RMS	30秒区間中の ストーム数のエントロピー	1	
		平均	1	
		中央値	1	
		標準偏差	1	
周波数 (11ビン) (0.1Hz幅ビン、0~1.1Hz)		11+1		
個々の RMS の大きさ (21ビン) (0G-20G)		21+1		
全体	規格化 ST	パワースペクトル密度 (5ビン) (1Hz幅ビン、0.5-5Hz)	5+1	24
		座位・歩行・走行のパーセンテージ	3	
		平均	1	
	中央値	1		
	標準偏差	1		
	周波数 (11ビン) (0.1Hz幅ビン、0~1.1Hz)	11+1		
パワースペクトル密度 (5ビン) (1Hz幅ビン、0.5-5Hz)	5+1			

2) ストレスレベル回帰分析プロセス

ストレスレベルの認識手法として、我々は先行研究のようなクラス分類ではなく回帰分析を選択した。中位を含む全範囲のストレスを認識するのに適切と判断したためである。図 4 を参照し、提案特徴量を使用した場合と使用しなかった場合のストレスレベル回帰分析プロセスを説明する。黒字 (赤字を含まない) は、提案特徴量を使用しないプロセスを示す。まず、33 人の被験者の生体信号について月単位で特徴量を計算した。その後特徴量の正規化を実施し、更に、L00CV (Leave One Out Cross Validation) ループを実行した。各 L00CV ループでは、1 人を抜いた 32 人の特徴量を訓練データ、PSS スコアを訓練ラベルとし、この訓練データ・訓練ラベルを用い、まず 10 を上限として特徴量選択を実施した。特徴量選択では特徴量と PSS スコアの相関係数をクライテリアとした。次に選択特徴量を用いて SVR (Support Vector Regression) によって抜いた 1 人の PSS スコアを推定し、この推定値を認識 PSS スコアとした。図 4 は、赤字を含めると、提案特徴量を使用した場合のプロセスを示す。特徴量は月単位だけでなく週単位でも計算された。この週単位の特徴量のプロットに対して最大値や標準偏差等の統計値が計算される。L00CV ループでは、提案特徴量セットと従来特徴量セットの両方から同じ数の特徴量を選択した。提案特徴量は主に中位のストレスレベル認識に有効であり、従来特徴量は上位/下位ストレスレベル認識に有効であるため、両方の特徴セットを使用することが全範囲のストレスレベルの認識には重要であるためである。

4.3 実験結果及び考察

1) 提案特徴量の有効性の評価

提案特徴量の有効性を評価するために、PSS スコアとの Pearson の相関係数が最も高い 1 つの提案特徴量に焦点を当てた。4 人の典型的な被験者の当該提案特徴量の週単位プロットを図 5 左に示す。一方、図 5 右の 3 つの棒グラフは、それぞれ、当該提案特徴量、PSS スコア、および当該提案特徴量に対応する従来特徴量の、4 被験者についての値を示している。ここでの従来特徴量は式 (2) の $a6G_M$ であり、提案特徴量は式 (4) の $a6G_{Wmax}$ である。継続的にストレスレベルの低い被験者 A および継続的にストレスレベルの高い被験者 D の週単位プロットは、それぞれ低い PSS スコア、高い PSS スコアをもたらすことは明らかであるが、これは PSS スコアの棒グラフで確認できる。当該棒グラフでは、被験者 A のスコアが最低であり、被験者 D のスコアが最高である。また、従来特徴量の棒グラフも、被験者 A、D については同じ傾向を示す。一方、週単位プロットが継続的に中程度である被験者 B と、第 1 週~第 3 週のプロットは低いものの第 4 週のプロットが高い被験者 C については、従来特徴量は PSS スコアと反対の傾向を示す (PSS スコアでは、被験者 C が被験者 B よりも高い一方、従来特徴量はその逆)。一方、提案特徴量は被験者 B、C についても PSS スコアと同じ傾向を示す。前節で指摘した通り、提案特徴量は PSS スコアの特性、つまり、特定の短い区間の強いストレスによって悪化した心理状態に影響を受けて高くなる、を的確に反映している。更に、これも前節で指摘した通り、この提案特徴量の有効性は特に中位 (図 5 においては被験者 B、C) のストレスレベルの認識において顕著である。

2) ストレスレベル認識性能

前節で述べたストレスレベル回帰分析プロセスでは、従来法 (従来特徴量のみ使用) と提案法 (従来特徴量と提案特

微量の双方を使用)の2つの方法において、上限10で特徴量選択を行い、選択された特徴量でSVRを行い、SVRによる認識PSSスコアと真のスコアとの間のPearson相関係数が最も高い場合を選択した。表2の「全体」列の、2つの上の行(「Pearsonの相関係数」と「誤差(RMSE)」)は、真のPSSスコアと認識PSSスコアとの間のPearsonの相関係数と誤差(RMSE(Root Mean Square Error))を示す。「全体」以外の2つの列(「低ストレス」および「中+高ストレス」)は、[8]で定義されている閾値(低ストレス:0-13, 中ストレス:14-26, 高ストレス:27-40)に基づいて真のPSSスコアにより低ストレスおよび中+高ストレスという二つのクラスに被験者を分けた場合の値を示している。高ストレスと中ストレスをひとつのクラスにまとめた理由は、用いたデータベースにおいて高ストレスクラスに2人の被験者しかおらず、しかもその2人のPSSスコアが閾値である27であったため、中ストレスクラスにまとめるべきだと判断したためである。「全体」列、上2行に示される値には、提案法の有効性が示されている。Pearsonの相関係数は0.66から0.72に向上し、RMSEが4.32から4.03に改善した。他の2つの列でも、提案法の使用による改善が示されている。

3) ストレスレベル順序認識性能

我々はまた、認識ストレスレベルの順序を用いて結果を評価した。例えば、図5の右側の棒グラフでは、4人の被験者の真のストレスレベルの順序は[A, B, C, D]である一方、従来特徴量による認識ストレスレベルの順序は、[A, C, B, D]である。この相違を調べることは、実用的な有用性を調査する上で重要である。なぜなら、実運用においては、ストレスレベルで並べた時、高い方にいる人は、相対的にリスクが高い人と判定する場合があるためである。前述

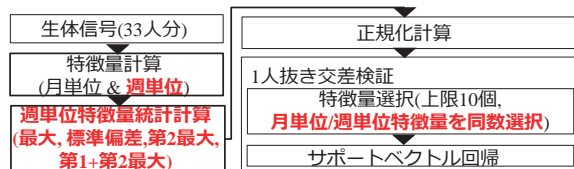


図4 PSSスコア回帰分析プロセス

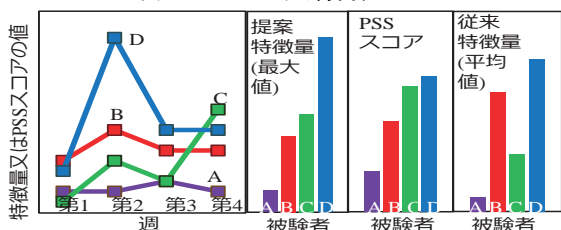


図5 例示特徴量の週単位プロット(左)と、提案法及び従来法による月単位計算結果及び比較のためのPSSスコア(右)

表2 提案法及び従来法のPearsonの相関係数, 誤差(RMSE), 平均序数差, 再現率-適合率カーブ下面積

		全体	低ストレス	中+高ストレス
Pearsonの相関係数	従来法	0.66	0.22	0.67
	提案法	0.72	0.37	0.70
誤差(RMSE)	従来法	4.32	6.24	3.50
	提案法	4.03	5.66	3.34
平均序数差	従来法	6.58	6.63	6.56
	提案法	5.45	5.63	5.40
再現率-適合率カーブ下面積	従来法	0.92		
	提案法	0.95		

の[A, B, C, D]と[A, C, B, D]の序数の差は、B, Cについてそれぞれの序数が2と3, 3と2で1ずつ異なるので、合計2となる。合計序数差を被験者数で割って、「平均序数差」も算出できる。提案法・従来法による認識PSSスコアと真のPSSスコアとの間の平均序数差を、表2の「平均序数差」行に示す。当該行において最大の改善は中+高ストレスクラスであり、四捨五入すれば7から5への改善である。PSSスコア分布(図3参照)では、頻度が7より大きいピンが1つしかない(該当するのはPSSスコア16-20のピンで、頻度は12)一方、PSSスコア14-27の中+高ストレス領域(PSSスコア11-15, 16-20, 21-25, 26-30の各ピンで包含される領域)では半分以上のピンの頻度が5より大きい。序数差と比較してピンの頻度が大きいということは、異なるピンつまりPSS分布の異なる位置と見なされる可能性が低いことを意味する為、これは顕著な改善であると言える。

4) ストレス識別性能

我々は中+高ストレスクラスを低ストレスクラスから識別する性能も評価した。我々の目的の一つは、中程度のストレスを感じている人を特定して高ストレスにならないよう注意を促す為だからである。低ストレスから中+高ストレスを識別するための再現率-適合率曲線のAUC(Area Under the Curve)を表2の最下段に示す。従来法の0.92に対し提案法は0.95となった。この顕著な改善(誤差削減幅約40%)は、提案法が、中程度のストレスの人を低ストレスの人から区別するのに有用であることを示している。

5. 結論

我々は長期ストレスレベルの認識においてより高い精度を達成する為、PSSがストレスレベルを認識する月単位だけでなく週単位で算出した特徴量も使用する方法を提案し、提案法を実験的に評価した。評価の為、オフィスワーカー33名の1ヶ月データを使用した。結果、提案法により様々な性能の改善が得られることが示された。例えば、Pearsonの相関係数は0.66から0.72に改善された。

参考文献

- [1] A. Alberdi, et al., "Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review," *Journal of biomedical informatics*, vol. 59, pp. 49-75, 2016.
- [2] A. Sano et al., "Recognizing academic performance, sleep quality, stress level, and mental health using personality traits, wearable sensors and mobile phones," in *Wearable and Implantable Body Sensor Networks (BSN)*, 2015 IEEE 12th International Conference on, 2015, pp. 1-6.
- [3] A. Sano, "Measuring College Students' Sleep, Stress, Mental Health and Wellbeing with Wearable Sensors and Mobile Phones," *Massachusetts Institute of Technology*, 2015.
- [4] S. Cohen, et al., "A Global Measure of Perceived Stress." *Journal of Health and Social Behaviour* 1983.
- [5] S. Cohen, et al., "Measuring Stress: A Guide for Health and Social Scientists." Oxford University Press, 1997.
- [6] <https://www.empatica.com/research/e4/>
- [7] S. Cohen and G. M. Williamson, "Perceived Stress in a Probability Sample in the United States", *The Social Psychology of Health*, 1988
- [8] State of New Hampshire Employee Assistance Program, "Perceived Stress Scale", URL: <https://das.nh.gov/wellness/Docs/Perceived%20Stress%20Scale.pdf>